

Kansas State University Libraries

**New Prairie Press**

---

Conference on Applied Statistics in Agriculture

2012 - 24th Annual Conference Proceedings

---

# TOWARDS BETTER FDR PROCEDURES FOR DISCRETE TEST STATISTICS

Xiongzhi Chen

R. W. Doerge

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

## Recommended Citation

Chen, Xiongzhi and Doerge, R. W. (2012). "TOWARDS BETTER FDR PROCEDURES FOR DISCRETE TEST STATISTICS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1040>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

# TOWARDS BETTER FDR PROCEDURES FOR DISCRETE TEST STATISTICS

Xiongzhi Chen and R.W. Doerge\*

Department of Statistics, Purdue University

**Abstract:** The false discovery rate (FDR) has been a widely used error measure in situations where a large number of tests are conducted simultaneously. Most methods that control the FDR at a prespecified level, or estimate the FDR of a multiple testing procedure (FDR procedures), were essentially developed for continuous test statistics. As such, their performances need to be carefully assessed when applied to discrete test statistics. We review some popular FDR procedures, point out a key reason for their excessive conservativeness when applied to discrete p-values, and suggest an improvement for these methods for such p-values.

**Keywords:** False discovery rate (FDR), Adaptive FDR procedures, Discrete p-values, Generalized FDR estimator

## 1 Introduction

A prominent feature of modern statistics is large-scale multiple testing, mainly driven by the field of genomics. For example, in differential gene expression studies the goal is to find relatively few differentially expressed genes among several thousand candidate genes under different biological conditions. In this example, a researcher has to simultaneously decide the status, differential expressed (DE) or none-differential expressed (NDE), for each gene. Inference on the status of each gene can be statistically thought of as a hypothesis testing problem, such that a true null means that a gene is NDE, whereas a false null means that a gene is DE. Interpreted this way, the task becomes one of multiple testing.

In typical multiple testing, there are  $m$  null hypotheses  $H_i$  with associated p-values  $p_i$ ,  $i = 1, \dots, m$ , for which  $H_i = 0$  means  $H_i$  is true (i.e., a true null) and  $H_i = 1$  means  $H_i$  is false (i.e., a false null or the alternative to  $H_i = 0$ ). Among these hypotheses,  $m_0$  are true nulls, and the rest false. The proportion of true nulls is defined as  $\pi = m_0/m$ . In practice,  $\pi \in (0, 1)$  but is unknown to the researcher. A p-value based multiple testing procedure (MTP)  $\mathcal{R}$  asserts each  $H_i = 0$  or 1 (i.e., accepts or rejects  $H_i$ ) based on its associated p-value  $p_i$ . It induces

$$\begin{cases} V(\mathcal{R}) = \# \text{ of rejected true nulls,} \\ S(\mathcal{R}) = \# \text{ of rejected false nulls,} \\ R(\mathcal{R}) = V(\mathcal{R}) + S(\mathcal{R}), \end{cases} \quad (1)$$

---

\*Corresponding author: R.W. Doerge, Department of Statistics, Purdue University, 250 N. University St., West Lafayette, IN 47907; email: doerge@purdue.edu; phone: +1-765-494-6030.

which are called the number of false discoveries, true discoveries, and total discoveries, respectively. When  $\mathcal{R}$  is based on some threshold  $t \in [0, 1]$  such that it asserts  $H_i = 0$  (or 1) according to  $p_i \geq t$  (or  $p_i < t$ ),  $\mathcal{R}$  in (1) is replaced by  $t$  and is called “one-step MTP” (Blanchard and Roquain, 2009).

For the gene expression example mentioned earlier, consider a naive MTP such that each hypothesis is tested at the same Type I error level  $\alpha \in (0, 1)$ . Regardless of whether there are any false nulls, we are expecting to have  $m\alpha$  false discoveries on average simply by chance. When  $m$  is large, this particular MTP may lead to many wrong inferential results, and is unfavorable. Alternatively, suppose we use the Bonferroni correction to control the family-wise error rate (FWER) at level  $\alpha$ , then  $H_i$  is rejected only when  $p_i < m^{-1}\alpha$ . When  $m$  is large, the rejection threshold  $m^{-1}\alpha$  can be so small that a false null can escape being rejected and discovered. Thus, such an MTP may detect much fewer false nulls than there actually are, and is undesirable. Further, in many applications, researchers are potentially able and willing to make more true findings by testing more hypotheses (e.g., by enlarging the pool of candidate genes in the study mentioned above) at the expense of a few more false discoveries. These considerations all call for a new MTP that dynamically balances the overall Type I error and statistical power even when the number of test increases drastically.

In response, Benjamini and Hochberg (1995) proposed a new concept called “the false discovery rate (FDR)” and an MTP that controls the FDR for p-values with continuous distributions. Following this seminal work, various adaptive FDR procedures, including those in Storey et al. (2004) and Benjamini et al. (2006) as the most popular, have been proposed. These procedures depend implicitly on the test statistics having continuous distributions and have been successfully applied to gene expression data from microarray technologies in the form of continuous measurements. However, with the next generation sequencing (NGS) technologies widely employed in gene expression studies, gene expression levels are recorded as discrete counts and researchers need to deal with statistical analysis of discrete data. In such endeavour, discrete p-values are induced by three routinely used tests: Fisher’s exact test (FET, Lehmann and Romano, 2005), the exact negative binomial test (ENBT, Robinson and Smyth, 2008), and permutation tests (Westfall and Young, 1993). Unlike the uniform distribution for p-values under true nulls in the continuous case, the p-value distributions induced by these tests are discrete and usually heterogeneous under true nulls. This presents a new phrase of multiple testing. Since almost all current FDR procedures incorporate an estimator of  $\pi$  that is designed for continuous p-values, a fundamental question arises: How efficient are current FDR procedures when applied to discrete p-values? Unfortunately, efforts in investigating the effects of discrete test statistics on FDR procedures and on developing FDR procedures for discrete test statistics are still insufficient. In this work, we review the most popular adaptive FDR procedures, point out their excessive conservativeness when applied to discrete p-values, and suggest new estimators that improve upon them.

## 2 Some popular FDR procedures

### 2.1 The vanilla Benjamini-Hochberg procedure

To potentially allow for more true discoveries when more discoveries are made, Benjamini and Hochberg (1995) proposed the concept of the discovery rate (FDR) for an MTP  $\mathcal{R}$  as,

$$FDR(\mathcal{R}) = E \left( \frac{V(\mathcal{R})}{R(\mathcal{R})} \middle| R(\mathcal{R}) > 0 \right),$$

where  $E$  is the expectation operator and the expectation is taken with respect to the joint distribution of the p-values. Since the FDR is the expected proportion of the number of false discoveries to that of total discoveries, controlling the FDR offers better balance between the overall Type I error and statistical power, and scales up naturally as the number of hypotheses to test increases. In order to control the FDR at a prespecified level  $\alpha \in (0, 1)$ , they also proposed a linear step-up MTP procedure for independent, continuous p-values (BH procedure), implemented as follows:

1. Sort the p-values  $p_i$  via their order statistics into  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$  and denote by  $H_{(i)}$  the hypothesis corresponding to  $p_{(i)}$ ,
2. Define step-up constants  $q_i = (i/m) \alpha$ ,  $i = 1, \dots, m$ , and

$$k_{BH} = \max \{ i : i \in \{ j : p_{(j)} \leq q_i \} \},$$

where  $\max \{ i : i \in \emptyset \} = -\infty$ ,

3. Reject  $H_{(i)}$  for which  $i \leq k_{BH}$ .

Benjamini and Hochberg (1995) showed that

$$FDR(p_{(k_{BH})}) = \pi \alpha \tag{2}$$

for independent p-values whose cumulative distribution functions (cdfs) under the true nulls (null cdfs) are uniform on the unit interval. The rejection threshold  $p_{(k_{BH})}$  of the BH procedure has an interesting interpretation. Suppose  $H_i$  is rejected if and only if  $p_i < \beta$  for some  $\beta \in (0, 1)$ , such that  $R_{m,\beta} > 0$  is the number of rejections it would have made, then the BH procedure chooses as rejection threshold  $\beta^*$  such that

$$\beta^* = \arg \max \left\{ \beta : \frac{\beta m}{R_{m,\beta}} \leq \alpha \right\}. \tag{3}$$

This reveals the intrinsic feature of BH procedure to maximize the total number of discoveries without exceeding the prespecified FDR level  $\alpha$ . Such a property of the BH procedure was further interpreted or extended in Storey et al. (2004) and Genovese and Wasserman

(2002, 2004). The BH procedure is indeed statistically more powerful than FWER control procedures in large-scale multiple testing as shown in Benjamini and Hochberg (1995) for p-values of continuous test statistics.

## 2.2 Adaptive BH procedure and Storey's FDR procedure

In extending the scope of application of the BH procedure, Benjamini and Yekutieli (2001) proved that

$$FDR_{BH} = FDR(p_{(k_{BH})}) \leq \pi\alpha, \quad (4)$$

for p-values, discrete or continuous, whereby the null distributions satisfy a certain type of positive dependency. We say that an FDR procedure is conservative at a prespecified FDR level  $\alpha$ , when its FDR is no less than  $\alpha$  or it provides an estimated FDR as an upper bound for  $\alpha$ . The result in (4) generalizes (2) to allow certain dependence among the p-values, and establishes the role of the proportion of true nulls in controlling the FDR of the BH procedure. Specifically, using a conservative estimator  $\hat{\pi}$  of the proportion of true nulls  $\pi$ , such that

$$\pi \leq \max\{\hat{\pi}, E[\hat{\pi}]\} < 1, \text{ with probability } 1 \quad (5)$$

when  $\pi < 1$ , gives a tighter FDR upper bound than using  $\pi = 1$ . With the same FDR level  $\alpha$ , this in turn yields a rejection threshold no smaller than  $p_{(k_{BH})}$  and leads to at least as many discoveries as the BH procedure. Such an observation, together with the fact that it is inappropriate to directly adjust  $m^{-1}R(\mathcal{R})$  when  $\mathcal{R}$  has been implemented and that  $\pi$  usually will appear as a factor in the FDR bound of an MTP, demonstrated that a “good” estimate of  $\pi$  in the sense of (5) holds the key to improved power of FDR procedures. It induces the adaptive BH procedure for which  $\alpha$  is replaced by  $\hat{\pi}^{-1}\alpha$  and

$$\hat{\pi}^{BH} = \frac{m - i + 1}{m(1 - p_{(i)})}, \quad 1 \leq i \leq m, \quad (6)$$

in Benjamini et al. (2006) can be used as  $\hat{\pi}$ . Further, it opened the door for adaptive FDR procedures that incorporate an estimate of  $\pi$  satisfying (5); see Storey et al. (2004), Benjamini et al. (2006), Blanchard and Roquain (2009) and others.

Among these adaptive procedures, a major conceptual extension of the BH procedure is Storey's FDR procedure in Storey et al. (2004):

$$\widehat{FDR}_{\lambda}(t) = \frac{\hat{\pi}^S(\lambda)t}{m^{-1} \max\{R(t), 1\}}, \quad t \in [0, 1], \quad (7)$$

where

$$\hat{\pi}^S(\lambda) = [(1 - \lambda)m]^{-1} \sum_{i=1}^m \mathbf{1}_{\{p_i > \lambda\}} \text{ for } \lambda \in [0, 1] \quad (8)$$

is an estimator of  $\pi$  with tuning parameter  $\lambda$  using indicator function  $\mathbf{1}_{\{\cdot\}}$ , and  $R(t) = \sum_{i=1}^m \mathbf{1}_{\{p_i \leq t\}}$  is the total number of discoveries.  $\widehat{FDR}_{\lambda=0}(t)$  is exactly the adaptive BH

procedure where  $\hat{\pi}$  is replaced by  $\hat{\pi}^S$  (Storey et al., 2004). Note that  $\widehat{FDR}_\lambda(t)$  is both an MTP using a threshold  $t$  and an estimator of its FDR. With

$$t_\alpha(\widehat{FDR}_\lambda) = \sup \left\{ t \in [0, 1] : \widehat{FDR}_\lambda(t) \leq \alpha \right\},$$

that is, the maximal rejection threshold for the p-values for which the estimated FDR of the MTP is controlled at level  $\alpha$ , Storey et al. (2004) justified that for independent or weakly dependent p-values with one extra assumption,  $\widehat{FDR}_\lambda$  provides conservative control and point estimation of the FDR of itself as an MTP using rejection threshold  $t_\alpha(\widehat{FDR}_\lambda)$ , that is,

$$FDR(\widehat{FDR}_\lambda(t_\alpha)) \leq \alpha. \quad (9)$$

The conservativeness of  $\widehat{FDR}_\lambda$  at  $t_\alpha(\widehat{FDR}_\lambda)$  in (9) endorses  $\widehat{FDR}_\lambda$  as an MTP with controlled FDR performance in many applications. (See Storey et al. (2004) for a variant of  $\hat{\pi}^S(\lambda)$  and of  $\widehat{FDR}_\lambda(t)$ , whose potential erroneous behavior is discussed in Chen and Dörge (2012).) However, it should be noted that  $\widehat{FDR}_\lambda(t)$  may display the so-called “lack of identifiability phenomenon” (LIP) found by Zhang et al. (2011), in the sense that for  $m$  given p-values, the minimal achievable level of the estimated FDR,  $\widehat{FDR}_\lambda(t)$ , in  $t$  for a fixed  $\hat{\pi}^S(\lambda)$  is greater than  $\alpha$ , i.e.,

$$\inf \left\{ \widehat{FDR}_\lambda(t) : t \in [0, 1] \right\} > \alpha.$$

In this case,  $\widehat{FDR}_\lambda$  fails to be a conservative FDR procedure, and should not be applied. Nonetheless, Storey’s FDR procedure, and the adaptive BH procedure, are the most well theoretically supported and widely used FDR procedures.

## 3 Adverse effects of discrete p-values

### 3.1 Excessive conservativeness of FDR procedures

Gilbert (2005) noticed that for discreteness of p-value distributions the FDR of the BH procedure can actually be smaller than  $\pi\alpha$  (c.f. (4)), and used the maximal value such a distribution can reach with respect to each step-up constant to derive a tighter FDR bound for the BH procedure for p-values of FET. Pounds and Cheng (2006) discussed the effects of discrete, non-uniformly distributed p-value distributions on estimation of  $\pi$  and on the FDR procedures, and showed that for permutation p-values, the BH procedure and the procedure in Storey (2002) are excessively conservative or erroneous. To mitigate this, they proposed

two estimators of  $\pi$  for two-sided discrete p-values as

$$\begin{cases} \hat{\pi}^{PC} = \min \left( 1, m^{-1} \sum_{i=1}^m \frac{p_i}{E[p_i | \text{ith null is true}]} \right), \\ \tilde{\pi}^{PC} = \min \left( 1, 2m^{-1} \sum_{i=1}^m p_i \right), \end{cases} \quad (10)$$

and demonstrated their improvement over  $\hat{\pi}^S(\lambda)$ . Estimators in (10) are the only published estimators of  $\pi$  that are built with the intention to accommodate discrete p-values, even though inefficiently as we will see next.

### 3.2 Excessive conservativeness of estimators of $\pi$

Note that (8) is deeply rooted in the assumption that a null p-value distribution is uniform and a relatively large p-value should have come from such a distribution, the expectation of  $\hat{\pi}^S(\lambda)$  can be much larger than  $\pi$  when applied to discrete p-values. The same is true for estimators in (10) since neither adjusts effectively for the discreteness of p-value distributions:  $\hat{\pi}^{PC}$  uses the average of the ratio of a realized p-value to its expectation under the true null and suffers from inflated summation caused by relatively large realized p-values; the discreteness of p-value distribution is ignored in the computation of  $\tilde{\pi}^{PC}$ . Therefore,  $\hat{\pi}^S$ ,  $\hat{\pi}^{PC}$  and  $\tilde{\pi}^{PC}$  are excessively conservative for discrete p-values (see Chen and Doerge (2012) for further details).

Finally,  $\hat{\pi}^{BH}$  in (6) is obtained by estimating  $m_0$  in three steps. First,  $i - 1$  p-values are rejected using rejection threshold  $p_{(i)}$  for some  $1 \leq i \leq m$ , which yields  $m - i + 1$  remaining hypotheses. Then

$$\hat{m}_0 = m - i + 1 + m_0 p_{(i)} \quad (11)$$

is used to approximate  $m_0$ , since among the  $i - 1$  rejected hypotheses an expected number of  $m_0 p_{(i)}$  false discoveries, i.e.,  $E(V(\mathcal{R})) = m_0 p_{(i)}$ , will have been made when the null p-values are uniformly distributed. Thirdly,  $\hat{\pi}^{BH} = \frac{\hat{m}_0}{m}$ , which is (6). From the previous description, we see that  $\hat{\pi}^{BH}$  also is based on the uniformity of a null p-value distribution, and that it can not perform well for discrete p-values. Clearly, whenever  $p_j < p_{(i)}$  for some  $1 \leq j \leq m$  and  $p_{(i)}$  is a point of discontinuity of the discrete cdf of  $p_j$ , rejecting  $H_j$  corresponding to  $p_j$  yields an expected “proportion” of false discovery strictly less than  $p_{(i)}$ , and

$$E(V(\mathcal{R})) < m p_{(i)} \quad (12)$$

holds. Since, when  $m$  is large, it is very unlikely that  $p_{(i)}$  is a point of continuity of all discrete p-values distributions, (12) is valid in general. Therefore, for discrete p-values,  $\hat{m}_0$ , in (11) is estimated to be larger than it should be, which consequently forces  $\hat{\pi}^{BH}$  to be excessively conservative. Similar behavior exhibits itself for all estimators of  $\pi$  that do not properly account for the discreteness of p-values. When an estimator  $\hat{\pi}$  of  $\pi$  is excessively conservative, so will be the adaptive FDR procedure that incorporates it.

### 3.3 Key reason for excessive conservativeness

To explicate the reason for this phenomenon, we present a graphical comparison of the uniform distribution of a p-value and a discrete p-value distribution under a true null in Figure 1. The graph depicts two properties of a discrete distribution. First, only finitely, or at most countably, many points of discontinuities can exist for a discrete cdf. These points form the support of a discrete p-value distribution, and the rest of a discrete p-value cdf is strictly below the diagonal line. This means that the discrete p-value cdf dominates the uniform distribution. Second, at each point outside the support of a discrete p-value cdf, there is a “gap” between the two distributions evaluated at that point. The message conveyed by Figure 1 is that, the more a discrete p-value distribution deviates from the uniform distribution, the larger the upward bias of an estimator of  $\pi$  is when it is designed for continuous p-values without taking into account the gaps.

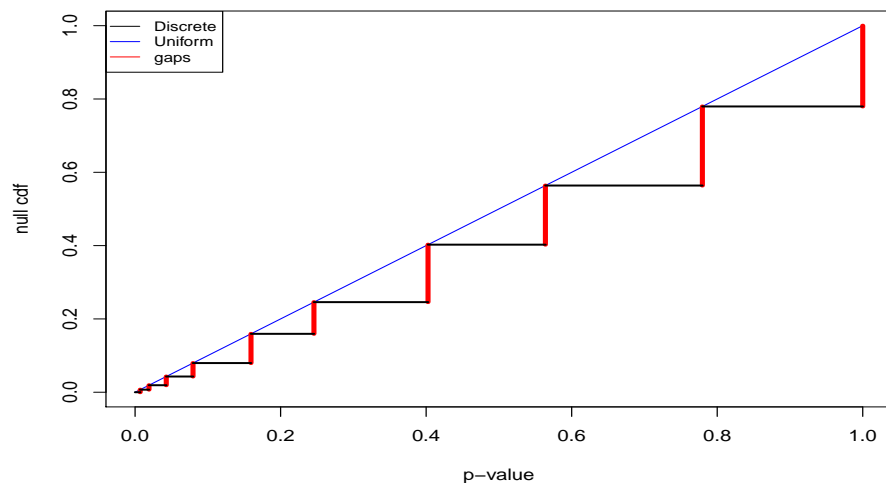


Figure 1: Null p-value distributions. The blue line is the uniform cdf, the red line segments together form the cdf of a null discrete p-value, and the red line segments are the differences between these two cdf’s at the support of the discrete p-value distribution.

## 4 New estimators of $\pi$ and FDR for discrete p-values

Since we understand the culprit for the over conservativeness of current estimators of  $\pi$  and the FDR, we suggest a new estimator  $\hat{\pi}^G$  of  $\pi$  by explicitly subtracting from  $\hat{\pi}^S(\lambda)$  in (8) the scaled difference between the discrete and uniform null p-value distributions  $F_i^*$  evaluated



at certain values in the range of a p-value. Specifically,

$$\hat{\pi}^G(\lambda, \epsilon) = \max \left[ 0, \min \left\{ 1, \hat{\pi}^S(\lambda) - \frac{\delta(\lambda, \epsilon)}{(1-\lambda)m} \right\} \right], \quad (13)$$

where

$$\delta(\lambda, \epsilon) = \sum_{i=1}^m \epsilon_i \{ \lambda - F_i^*(\lambda) \},$$

and  $\epsilon = (\epsilon_1, \dots, \epsilon_m)$  is a vector of pre-specified or adaptively estimated constants (usually in  $[0, 1]$ ). In the term  $\delta(\lambda, \epsilon)$ , each  $\lambda - F_i^*(\lambda)$  is the difference between the uniform distribution and the discrete cdf of p-value  $p_i$  evaluated at  $\lambda$ , i.e., the gap between the two distributions at  $\lambda$ . This gap is then scaled by  $\epsilon_i$  to additionally adjust for the heterogeneity of null discrete p-value distributions, and account for the existence of false nulls. Finally, all such gaps are summed up into  $\delta(\lambda, \epsilon)$ , which is then scaled by  $(1-\lambda)^{-1}m^{-1}$  to be technically compatible with  $\hat{\pi}^S$  and subtracted from  $\hat{\pi}^S$ . Such a design makes  $\hat{\pi}^G$  considerably less conservative than its counterparts.

The new estimator  $\hat{\pi}^G$  induces a new adaptive FDR estimator  $\widetilde{FDR}$  as

$$\widetilde{FDR}_\lambda(t) = \min \left[ 1, \frac{\hat{\pi}^G(\lambda, \epsilon)t}{m^{-1} \max\{R(t), 1\}} \right].$$

By its definition,  $\widetilde{FDR}$  is less conservative than  $\widetilde{FDR}_\lambda$  and the adaptive BH procedure. We call  $\hat{\pi}^G$  and  $\widetilde{FDR}$  generalized estimators of  $\pi$  and of the FDR, respectively. With properly chosen  $\epsilon$ ,  $\hat{\pi}^G$  and  $\widetilde{FDR}$  remove all defects of  $\hat{\pi}^S$  and  $\widetilde{FDR}_\lambda$  except the LIP because, as  $\widetilde{FDR}_\lambda$ ,  $\widetilde{FDR}$  is both an MTP and FDR estimator.

For continuous p-values,  $\hat{\pi}^G$  reduces to  $\hat{\pi}^S(\lambda)$  (since each  $\lambda - F_i^*(\lambda)$  equals zero), and so does  $\widetilde{FDR}_\lambda$  to  $\widetilde{FDR}_\lambda$ . Hence,  $\hat{\pi}^S$  and  $\widetilde{FDR}_\lambda$  inherit all excellent properties of Storey's estimators in this case. Yet, by definition they are (considerably) less conservative in estimating  $\pi$  and the FDR when applied to discrete p-values. The generalized estimators are to our knowledge the only existing estimators that effectively adjust for the discreteness of p-values, and are applicable, with assured performance, to both continuous and discrete p-values. In Chen and Doerge (2012), we have both theoretically and empirically showed that  $\hat{\pi}^G(\lambda, \epsilon)$  and  $\widetilde{FDR}_\lambda$  outperform their competitors for discrete p-values, and perform equally well as the leading estimators of these two quantities for continuous p-values.

## 5 Summary

We have restated the importance of estimators of the proportion of true nulls to improve the power of MTPs, reviewed two of the most popular adaptive FDR procedures, and pointed out their excessive upward bias when estimating  $\pi$  and the FDR for discrete test statistics in

terms of p-values. We have clarified a key reason for the over conservativeness of current FDR procedures for discrete p-values, and suggested new estimators of  $\pi$  and of the FDR that are applicable, with excellent performance, to p-values arising from discrete and continuous data.

At the time of writing this paper, we are aware of the most recent work, Heller and Gur (2012), on FDR control procedures for discrete test statistics, where “mid-p-values” are used in place of the original discrete p-values. Their method and the one presented here are complementary to each other, but with a key difference, in that ours keeps the derivation process of a discrete p-value intact. We are anticipating further improvement in the estimation and control of the FDR when these two methods are properly integrated and applied to discrete test statistics.

The work of Heller and Gur (2012) relates to a fundamental issue on how statistical inference with discrete test statistics are made in practice. It is widely acknowledged that for various exact test procedures, which are themselves the uniformly most powerful (UMP) (Lehmann and Romano, 2005), the corresponding two-sided p-values are not computed based on the acceptance and rejection regions of these tests under a given size but in a less efficient way as done, for example, in Agresti (2002). This warns us of a possible loss in power when statistical decisions are based on such p-values, and advocates estimating  $\pi$  and the FDR from the test statistics directly.

Besides the two major issues of multiple testing for discrete test statistics, namely, discreteness of their distribution, and how p-values are derived for such test statistics, there are certainly other facets that we should consider and explore. Driven by the quick accumulation of massive discrete data together with accompanying large-scale statistical inference, the need for better FDR procedures for this type of data is urgent. We hope that the key reason for the over-conservativeness of popular FDR procedures for discrete p-values, and their intrinsic limits even when applied to continuous test statistics discussed here will stimulate further research in multiple testing. Moreover, we expect that the generalized estimators suggested here and fully studied in our technical report Chen and Doerge (2012) will serve as a better alternative for FDR control and estimation for discrete p-values.

## References

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Inc., New Jersey.
- Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57(1), 289–300.
- Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3), 491–507.
- Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 29(4), 1165–1188.

- Blanchard, G. and E. Roquain (2009). Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research* 10, 2837–2871.
- Chen, X. and R. Doerge (2012). Generalized estimators of the proportion of true nulls and of the false discovery rate for multiple testing. *Technical report. Department of statistics, Purdue University, West Lafayette.*
- Genovese, C. and L. Wasserman (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Statist. Soc. B* 64(3), 499–517.
- Genovese, C. and L. Wasserman (2004). A stochastic process approach to false discovery control. *Ann. Statist.* 32(3), 1035–1061.
- Gilbert, P. B. (2005). A modified false discovery rate multiple-comparisons procedure for discrete data, applied to human immunodeficiency virus genetics. *J. R. Statist. Soc. C* 54(1), 143–158.
- Heller, R. and H. Gur (2012). False discovery rate controlling procedures for discrete tests. <http://arxiv.org/abs/1112.4627>.
- Lehmann, E. and J. P. Romano (2005). *Testing Statistical Hypotheses*. Springer, New York.
- Pounds, S. and C. Cheng (2006). Robust estimation of the false discovery rate. *Bioinformatics* 22(16), 1979–1987.
- Robinson, M. D. and G. K. Smyth (2008). Small-sample estimation of negative binomial dispersion, with applications to sage data. *Biostatistics* 9(2), 321–332.
- Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc. B* 64(3), 479–498.
- Storey, J. D., J. E. Taylor, and D. Siegmund (2004). Strong control, conservative point estimation in simultaneous conservative consistency of false discover rates: a unified approach. *J. R. Statist. Soc. B* 66(1), 187–205.
- Westfall, P. and S. Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley-Interscience, New York.
- Zhang, C., J. Fan, and T. Yu (2011). Multiple testing via  $\text{fdr}_l$  for large-scale imaging data. *Ann. Statist.* 39(1), 613–642.